**University of São Paulo
“Luiz de Queiroz” College of Agriculture**

**Predicting the performance of untested maize single cross hybrids
based on information from genomic relationship matrix and genotype
by environment interaction**

Matheus Dalsente Krause

Dissertation presented to obtain the degree of Master in
Science. Area: Genetics and Plant Breeding

**Piracicaba
2018**

Matheus Dalsente Krause
Agronomist

**Predicting the performance of untested maize single cross hybrids
based on information from genomic relationship matrix and genotype
by environment interaction**

Advisor:

Prof. Dr. **ANTONIO AUGUSTO FRANCO GARCIA**

Dissertation presented to obtain the degree of Master in
Science. Area: Genetics and Plant Breeding

Piracicaba
2018

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP

Krause , Matheus Dalsente

Predicting the performance of untested maize single cross hybrids based on information from genomic relationship matrix and genotype by environment interaction / Matheus Dalsente Krause . -- Piracicaba, 2018 .

63 p.

Dissertação (Mestrado) -- USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Seleção genômica 2. GBLUP 3. Ensaios para múltiplos ambientes 4. Variância-covariância I. Título.

DEDICATORY

To my wife, Aline.

ACKNOWLEDGMENTS

- To the “Luiz de Queiroz” College of Agriculture - University of São Paulo (ESALQ/USP) for all support to develop my Master’s Degree and this project.
- To the Coordination for the Improvement of Higher Education Personnel (CAPES) for my scholarship.
- To my advisor Prof. Dr. Antonio Augusto Franco Garcia for believing in my potential as a graduate student.
- To Embrapa Maize and Sorghum, an important partner of this project.
- All students (and friends) in the Statistical Genetics Laboratory at ESALQ/USP. We have made a great team!
- To all professors and staff of Genetics Department and Exact Sciences Department at ESALQ/USP.
- To my parents for their support.

Thanks.

SUMMARY

Resumo	6
Abstract	7
1 Introduction	9
References	9
2 Conclusions	13

RESUMO

Predição de híbridos simples de milho não avaliados com informações da matriz de parentesco realizada e interação genótipos por ambientes

A fenotipagem em ensaios de múltiplos ambientes (MET) tem papel importante para acessar a resposta diferencial de híbridos de milho em diferentes regiões alvo de melhoramento, o que se deve a interação genótipos por ambientes (GxE). Neste contexto, um modelo efetivo de seleção genômica (GS) para predição do desempenho de híbridos não avaliados em MET é essencial para maximizar os ganhos genéticos e alocar eficientemente o orçamento dos programas de melhoramento. Desta forma, os objetivos deste estudo foram (i) avaliar as acurácias preditivas de modelos GBLUP (do inglês, *Genomic Best Linear Unbiased Prediction*) na predição da produtividade de grãos de híbridos simples de milho tropical não avaliados, usando modelos genético-estatísticos que levam em consideração a interação GxE através de uma estrutura de variância-covariância (VCOV) do tipo fator analítico (FA) e (ii) investigar a utilidade da matriz de parentesco realizada em combinação com diferentes estruturas de VCOV para efeitos genéticos e de resíduos em diferentes níveis de ambientes em desbalanceamento. As predições foram realizadas em duas situações: (CV1) híbridos não avaliados em nenhum ambiente e (CV2) híbridos avaliados em alguns ambientes e em outros não. Foram fenotipados 156 híbridos simples de milho em 12 ambientes para a característica produtividade de grãos. O genótipo dos híbridos foi inferido com base nas informações de marcadores SNP (do inglês, *single nucleotide polymorphism*) das linhagens parentais, obtidos via GBS (do inglês, *genotyping-by-sequencing*). Os procedimentos e modelos utilizados neste estudo podem ser facilmente estendidos a outras culturas em que MET desempenha um papel importante no processo de melhoramento.

Palavras-chave: Seleção Genômica; GBLUP; Ensaios para Múltiplos Ambientes; Variância-Covariância

ABSTRACT

Predicting the performance of untested maize single cross hybrids based on information from genomic relationship matrix and genotype by environment interaction

Phenotyping in multi-environment trials (MET) plays an important role to access the differential response of maize hybrids across target breeding regions due to genotype by environment (GxE) interaction. In this context, an effective model of genomic selection (GS) to predict the performance of untested hybrids in MET is essential to maximize genetic gains and to efficiently allocated the breeding programs' budget. Therefore, the goals of this study were (*i*) to evaluate the predictive accuracies of GBLUP (Genomic Best Linear Unbiased Prediction) models to predict grain yield performance of unobserved tropical maize single-cross hybrids, using models that consider GxE interaction by fitting a factor analytic (FA) variance-covariance (VCOV) structure, and (*ii*) to investigate the usefulness of genomic relationship information in combination with different VCOV for genetics and residuals effects, under different levels of unbalanced environments. Predictions were performed for two situations: (CV1) untested hybrids, and (CV2) hybrids evaluated in some environments but missing in others. Phenotypic data of grain yield was measured in 156 maize single-cross hybrids at 12 environments. Hybrids genotypes were inferred based on their parents (inbred lines) via SNP (single nucleotide polymorphism) markers obtained from GBS (genotyping-by-sequencing). The procedures and models applied in this study can be easily extended to other crops in which MET plays an important role in the breeding process.

Keywords: Genomic Selection; GBLUP; Multi-Environment Trials; Variance-Covariance

1 INTRODUCTION

Maize (*Zea mays* L.) plays an important role in global food security, being a key crop for 460 million inhabitants in sub-Saharan Africa, Asia, and Latin America (PRASANNA, 2016, p.62). Worldwide, in terms of animal supply chain, 60-70% of harvested maize is used as livestock feed (GWIRTZ and GARCIA-CASAL, 2014). It is also has been important for biofuel production (SHIFERAW *et al.*, 2011). By 2050 the expected population on Earth is 9.8 billion people (UN DESA, 2017), and whereas maize yields remain low in many developing countries, Brazil can be an important player to feed the world. Brazil already is a big maize producer ranking in the third position globally (USDA, 2017), and to properly face the task, the challenge to release fast superior cultivars to the market has intensified.

Genomic Selection (GS) has been recently incorporated into plant breeding programs (JONAS and DE KONING, 2016) and is a promising tool to predict unobserved maize single-cross hybrids (BURGUEÑO *et al.*, 2012; CROSSA *et al.*, 2017). Proposed by MEUWISSEN *et al.* (2001), it consists in predicting the genetic merit of a genotype based on molecular markers information covering the whole genome. The availability of large-scale genomic information for most crops, due to cost-effective high-throughput sequencing technologies, is an important contributor for the success of GS (CROSSA *et al.*, 2017). Molecular markers information can be used to estimate the identity-by-state relationship between pair of individuals (VANRADEN, 2008; POWELL *et al.*, 2010) and this information can be accounted by linear mixed models based on the genomic realized relationship matrix $\mathbf{A}$.

One of the most resource-demanding phase in a breeding program consists of hybrids phenotyping in multi-environment trials (MET) (FRITSCHÉ-NETO *et al.*, 2010). MET are crucial to access hybrids performance, allowing breeders to quantify the differential response of hybrids across target breeding regions or environments, phenomena known as genotype-by-environment (GxE) interaction. High levels of predictive accuracies have been found for GS models that incorporate both genomic information and GxE interaction (JARQUÍN *et al.*, 2014; ACOSTA-PECH *et al.*, 2017). Data from MET are usually unbalanced due to the natural process of selection; hybrids with poor performance are discarded and new entries are added every year (PIEPHO *et al.*, 2008; DAWSON *et al.*, 2013). Therefore, GS models that properly deal with unbalanced data in MET leads to better predictive accuracies across environments.

The implementation of GS into breeding programs reshaped the breeder's equation; rather than genetic gains per cycle of selection (generation interval), gains per unit time/annual rate have been taken into account and more efficiently allocated budget in the breeding program (HEFFNER *et al.*, 2010; HICKEY *et al.*, 2012). Thus, this dissertation have the goal of evaluating the flexibility of linear mixed models to account complex variance-covariance structures, considering genomic and MET information, under different levels of missing data across environments.

References

ACOSTA-PECH, R., J. CROSSA, G. DE LOS CAMPOS, S. TEYSSÈDRE, B. CLAUSTRES, S. PÉREZ-ELIZALDE, and P. PÉREZ-RODRÍGUEZ, 2017 Genomic models with genotype $\times$

- environment interaction for predicting hybrid performance: an application in maize hybrids. *Theoretical and Applied Genetics* **130**: 1431–1440.
- BURGUEÑO, J., G. DE LOS CAMPOS, K. WEIGEL, and J. CROSSA, 2012 Genomic prediction of breeding values when modeling genotype $\times$ environment interaction using pedigree and dense molecular markers. *Crop Science* **52**: 707–719.
- CROSSA, J., P. PÉREZ-RODRÍGUEZ, J. CUEVAS, O. MONTESINOS-LÓPEZ, D. JARQUÍN, G. DE LOS CAMPOS, J. BURGUEÑO, J. M. GONZÁLEZ-CAMACHO, S. PÉREZ-ELIZALDE, Y. BEYENE, S. DREISIGACKER, R. SINGH, X. ZHANG, M. GOWDA, M. ROORKIWAL, J. RUTKOSKI, and R. K. VARSHNEY, 2017 Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science* **22**: 961–975.
- DAWSON, J. C., J. B. ENDELMAN, N. HESLOT, J. CROSSA, J. POLAND, S. DREISIGACKER, Y. MANÈS, M. E. SORRELLS, and J. L. JANNINK, 2013 The use of unbalanced historical data for genomic selection in an international wheat breeding program. *Field Crops Research* **154**: 12–22.
- FRITSCHÉ-NETO, R., M. C. GONÇALVES, R. VENCovsky, and C. L. D. SOUZA JUNIOR, 2010 Prediction of genotypic values of maize hybrids in unbalanced experiments. *Crop Breeding and Applied Biotechnology* **10**: 32–39.
- GWIRTZ, J. A. and M. N. GARCIA-CASAL, 2014 Processing maize flour and corn meal food products. *Annals of the New York Academy of Sciences* **1312**: 66–75.
- HEFFNER, E. L., A. J. LORENZ, J. L. JANNINK, and M. E. SORRELLS, 2010 Plant breeding with Genomic selection: Gain per unit time and cost. *Crop Science* **50**: 1681–1690.
- HICKEY, J. M., J. CROSSA, R. BABU, and G. DE LOS CAMPOS, 2012 Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. *Crop Science* **52**: 654–663.
- JARQUÍN, D., J. CROSSA, X. LACAZE, P. DU CHEYRON, J. DAUCOURT, J. LORGEU, F. PIRAUX, L. GUERREIRO, P. PÉREZ, M. CALUS, J. BURGUEÑO, and G. DE LOS CAMPOS, 2014 A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theoretical and Applied Genetics* **127**: 595–607.
- JONAS, E. and D. J. DE KONING, 2016 Goals and hurdles for a successful implementation of genomic selection in breeding programme for selected annual and perennial crops. *Biotechnology and Genetic Engineering Reviews* **32**: 18–42.
- MEUWISSEN, T. H. E., B. J. HAYES, and M. E. GODDARD, 2001 Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**: 1819–1829.
- PIEPHO, H. P., J. MÖHRING, A. E. MELCHINGER, and A. BÜCHSE, 2008 BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* **161**: 209–228.
- POWELL, J. E., P. M. VISSCHER, and M. E. GODDARD, 2010 Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Reviews Genetics* **11**: 800–805.

- PRASANNA, B., 2016 Developing and Deploying Abiotic Stress-Tolerant Maize Varieties in the Tropics: Challenges and Opportunities. In *Molecular Breeding for Sustainable Crop Improvement*, edited by V. R. Rajpal, S. R. Rao, and S. Raina, volume 11, chapter 3, pp. 61–77.
- SHIFERAW, B., B. M. PRASANNA, J. HELLIN, and M. BÄNZIGER, 2011 Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Security* **3**: 307–327.
- UN DESA, 2017 World population projected to reach 9.8 billion in 2050, and 11.2 billion in 2100.
- USDA, 2017 World Agricultural Supply and Demand Estimates.
- VANRADEN, P. M., 2008 Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**: 4414–4423.

2 CONCLUSIONS

(i) The inclusion of factor analytic structure boosted the predictive accuracy of untested maize single-cross hybrids, regardless residuals modeling;

(ii) Models that included genomic relationship information and GxE interaction by factor analytic structure achieved higher predictive accuracy in elevated levels of missing environments; and

(iii) High levels of predictive accuracy of untested maize single-cross hybrids were found with moderated to low levels of missing environments.